

Centralne repozytorium wiedzy Politechniki Poznańskiej - koncepcja, metodologia, organizacja.

Beata Korzystka¹, Krzysztof Ober¹

¹Politechnika Poznańska, Biblioteka, ul. Piotrowo 2, 61-138 Poznań
{beata.korzystka, krzysztof.ober}@put.poznan.pl

1. Wstęp

Dokumentowanie i rozpowszechnianie informacji o dorobku naukowym uczelni to jedno z priorytetowych działań współczesnych bibliotek akademickich. Biblioteki te, jako że są zintegrowane ze strukturą szkoły wyższej, starają się włączać w realizację celów stawianych obecnie polskim uczelniom poprzez wdrażanie nowoczesnych systemów. Systemy przeznaczone do wspierania powyższych zadań są dzisiaj nieodzowne w funkcjonowaniu uczelni wyższej. Kluczowym elementem w podjęciu decyzji o wdrożeniu nowego systemu jest położenie nacisku (w ramach długofalowych działań) na jego elastyczność, otwartość, łatwość modyfikacji i możliwości dostosowywania go do bieżących potrzeb. Należy również zauważyć, że system taki stanowi duże wsparcie w pracy naukowej indywidualnych pracowników oraz jednostek uczelni, a poprzez możliwość generowania różnych zestawień i raportów, jest niezbędnym narzędziem pracy przy corocznej ocenie, kategoryzacji oraz w innych sytuacjach wymagających zestawień dorobku naukowego.

Z systemu korzystać będą pracownicy naukowcy, doktoranci, studenci, władze uczelni i pracownicy administracji oraz bibliotekarze.

Obecnie Biblioteka Politechniki Poznańskiej dysponuje dwoma systemami (baza bibliografii BIBLIO i Repozytorium Naukowe PP), których funkcjonalności tylko częściowo spełniają oczekiwania docelowych grup użytkowników. Zasadniczą ich wadą jest to, że funkcjonują oddzielnie i na etapie ich tworzenia były przeznaczone do odrębnych celów. Baza Biblio służy do gromadzenia informacji bibliograficznej o dorobku naukowym, a Repozytorium do dokumentowania dorobku naukowego w wersji pełnotekstowej.

Zintegrowanie zawartości dwóch obecnie eksploatowanych systemów, a także pozyskanie informacji o dorobku naukowym z innych źródeł, w formie centralnego repozytorium wiedzy jest dla funkcjonowania i rozwoju uczelni sprawą priorytetową, gdyż funkcjonalności nowego systemu po pierwsze spełniają aktualne wymagania MNiSW w zakresie oceny i parametryzacji, a po drugie wspierają wiele procesów decyzyjnych uczelni oraz promują i eksponują w jednolitej formie cenny dorobek naukowy w globalnym świecie nauki. Celowość podjęcia powyższych działań i uzasadnienie ich priorytetowości zostało poparte wnioskiem do MNiSW o sfinansowanie powyższego projektu dotyczącego utworzenia centralnego repozytorium wiedzy Politechniki Poznańskiej ze środków na Działalność Upowszechniającą Naukę.

2. System

Centralne repozytorium wiedzy Politechniki Poznańskiej powstaje w oparciu o oprogramowanie OMEGA-PSIR (Ω - Ψ R) (<http://omegapsir.ii.pw.edu.pl/>) opracowane w Instytucie Informatyki Politechniki Warszawskiej w ramach projektu Synat (<http://www.synat.pl/>).

System ten zaimplementowano w technologii Java z wykorzystaniem repozytorium treści Apache Jackrabbit spełniającym założenia standardu Java Content Repository (*JCR, specyfikacje JSR 170 i JSR 283*). Dzięki temu możliwe jest przechowywanie obiektów cyfrowych i zarządzanie nimi, w tym m. in.: indeksowanie, wersjonowanie i obserwację zmian w treści, kontrolę uprawnień, a także łatwą integrację na poziomie danych oraz swobodną zmianę dostawcy implementacji.

System przechowuje informacje o obiektach różnego typu, takich jak afiliacja, autor, konferencja, czasopismo, publikacja. Istnieje możliwość uzupełniania danych w systemie w drodze importu.

Zaimplementowany został mechanizm pełnego importu w formacie XML dla danych każdego typu oraz importu w formacie BibTeX dla danych typu publikacja (artykuł, książka, rozdział, raport, doktorat).

Struktura plików XML musi odpowiadać strukturze danych przyjętej dla poszczególnych typów obiektów opisanej za pomocą XSD (*XML Schema Definition*) oraz we właściwy sposób musi odzwierciedlać relacje pomiędzy obiektami i ich hierarchię.

Dzięki dostępności kodów źródłowych systemu oraz zastosowaniu standardu XSD możliwa jest odpowiednia modyfikacja typów i struktur danych w celu dostosowania ich do własnych potrzeb. Istnieje też możliwość ograniczonej modyfikacji struktury danych na etapie wprowadzania danych do systemu poprzez mechanizm pól dodatkowych (*userfields*) (np. dodanie pól związanych z prawami i licencją w przypadku rozpraw doktorskich) (rys. 1, rys. 3).

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<collection>
  <phd>
    <author><name>Marcin</name><surname>Pelic</surname></author>
    <titlePL>Badanie nieliniowości belkowego przetwornika piezoelektrycznego</titlePL>
    <titleEN>Investigation of the nonlinearity of piezoelectric transducer</titleEN>
    <supervisor>...</supervisor>
    <certifyingUnit>...</certifyingUnit>
    <externalreviewers>...</externalreviewers>
    <classOPI>budowa i eksploatacja maszyn</classOPI>
    <classUE>Inżynieria mechaniczna</classUE>
    <discipline>...</discipline>
    <abstractPL>...</abstractPL>
    <abstractEN>...</abstractEN>
    <keywordsPL>przetwornik piezoelektryczny ; histereza ; kompensacja</keywordsPL>
    <keywordsEN>piezoelectric transducer ; hysteresis ; compensation</keywordsEN>
    <userfield>
      <key>Uzyskany tytuł</key>
      <value>doktor nauk technicznych w dyscyplinie: budowa i eksploatacja maszyn</value>
    </userfield>
    <userfield>
      <key>Licencja</key>
      <value>udostępnianie na podstawie umowy z właścicielem majątkowych praw autorskich</value>
    </userfield>
    <userfield>
      <key>Właściciel praw</key>
      <value>Pelic, Marcin Tadeusz</value>
    </userfield>
    <userfield>
      <key>Lokalizacja oryginału</key>
      <value>BPP</value>
    </userfield>
    <userfield>
      <key>Prawa do dysponowania publikacją</key>
      <value>Politechnika Poznańska</value>
    </userfield>
    <userfield>
      <key>Prawa</key>
      <value>wszystkie prawa zastrzeżone</value>
    </userfield>
    <userfield>
      <key>Prawa dostępu</key>
      <value>dla wszystkich w zakresie dozwolonego użytku</value>
    </userfield>
  </phd>
</collection>
```

Rys. 1. Rozszerzanie struktury danych za pomocą pól dodatkowych – widok XML

System OMEGA-PSIR jest dostarczany w postaci zestawu plików, który należy umieścić w odpowiedniej strukturze katalogów na serwerze. Szczegółowa instrukcja zawiera wszelkie informacje dotyczące wymagań sprzętowych jak i programowych, w tym również dokładny opis sposobu przygotowania środowiska systemowego oraz wyszczególnienie działań, które należy podjąć w celu poprawnej konfiguracji i uruchomienia systemu repozytorium.

Pakiet instalacyjny zawiera także zestaw danych systemowych i walidacyjnych:

- algorytmy punktacji,
 - algorytmy oceny eksperta
- oraz zestaw innych danych wspierających korzystanie z aplikacji, m. in.:
- baza czasopism,
 - baza konferencji (wykaz tytułów i wykaz wydarzeń),
 - baza instytucji,
 - języki,
 - typy projektów.

Po utworzeniu struktury jednostek uczelni i załadowaniu danych pracowników (z wykorzystaniem mechanizmu importu XML) można rozpocząć wprowadzanie (importowanie) danych publikacji.

3. Analiza źródeł informacji

Działania zmierzające do przygotowania metadanych do importu wymagały wielu analiz i podejść rozłożonych na kilka etapów, w tym pozyskanie jednoznacznej informacji w sposób jak najbardziej zautomatyzowany, z wykorzystaniem funkcjonalności dostarczanych przez poszczególne platformy, jak również wspieranie ich własnymi rozwiązaniami (zestawy skryptów w języku Perl do pobierania, konwersji, scalania i weryfikacji danych). Mając na uwadze, że ogromna liczba opisów bibliograficznych publikacji naukowych, które docelowo powinny zostać zaindeksowane w centralnym systemie, jest mocno rozproszona (w różnych źródłach: bazach danych, serwisach czasopism pełnotekstowych, serwisach książek, katalogach bibliotecznych, platformach wydawniczych, czy różnych stronach WWW - Google Scholar, Google Patents itp.), podjęto działania, które pozwolą na zebranie ich w sposób uporządkowany za pomocą różnego rodzaju identyfikatorów (indywidualnych numerów w bazie identyfikujących jednoznacznie autora i jego publikacje). Ogromnym źródłem informacji o dorobku naukowym naszych pracowników, które zapewnia pozyskanie takiego numeru, a wraz z nim listy publikacji, jest baza Scopus. Drugim równie ważnym, a tak naprawdę z punktu widzenia MNiSW najważniejszym źródłem informacji o dorobku naukowym jest baza Web of Science, która również pozwala na jednoznaczną identyfikację autora i jego publikacji, ale wymaga to podjęcia pewnych dodatkowych działań. W tym celu zostało opracowane odpowiednie zarządzenie, (*Zarządzenie nr 13 Rektora Politechniki Poznańskiej z dnia 8 maja 2014 r. RO/V/13/2014 w sprawie uzyskanie indywidualnego numeru Researcher ID w bazie Web of Science*), które dzięki zaangażowaniu pracowników naukowych pozwala na sukcesywne gromadzenie informacji o publikacjach indeksowanych w bazie Web of Science (WoS) na profilach ResearcherID, w tym również publikacji nie posiadających afiliacji jednostki macierzystej. Podejście takie wynikało z faktu, że wyszukiwanie publikacji w bazie WoS poprzez opcję afiliacji jednostki nie pozwala na wygenerowanie pełnego dorobku autora i w efekcie wynik wyszukiwania jest zaniżony. Lista reluzatów nie zawiera publikacji autora stworzonych między innymi w ramach współpracy z przemysłem, czy zagranicznego stażu. Ponadto funkcjonalność profilu ResearcherID, pozwala również na dodawanie informacji o pozostałym dorobku naukowym, który nie jest indeksowany w Web of Science. Wśród pracowników naukowych rozwinęła się również inicjatywa tworzenia profili Google Scholar, które także służą do gromadzenia dorobku. Wykorzystanie jednoznacznej identyfikacji poprzez unikatowe numery ID autora i nazwy profili Google Scholar ma posłużyć wstępnemu zebraniu możliwie największej ilości danych, ich uporządkowaniu i sukcesywnemu wprowadzaniu do docelowego systemu.

Na dalszym etapie zbierania informacji o publikacjach przeanalizowano możliwości pozyskiwania danych i przeprowadzono przykładowe eksporty danych opisów bibliograficznych z następujących źródeł informacji:

- Web of Science (<http://apps.webofknowledge.com/>),
- Scopus (<http://www.scopus.com/>),
- Google Scholar (<http://scholar.google.pl/>),
- SciFinder - Chemical Abstract (<https://scifinder.cas.org/>),
- Science Direct (<http://www.sciencedirect.com/>),
- BazTech (<http://baztech.icm.edu.pl/>),
- Bibliografia Publikacji Pracowników PP– BIBLIO (<http://library.put.poznan.pl/bib/index.html>),
- Repozytorium Naukowe Politechniki Poznańskiej (<http://repozytorium.put.poznan.pl>)
- bazy danych i serwisy obejmujące dokumentację patentową, o której szerzej napisano w dalszej części artykułu poświęconej wybranym typom publikacji (patrz. p. 5).

Wzięto pod uwagę fakt, że baza BIBLIO jest bazą bibliograficzną dokumentującą dorobek pracowników PP od 1994 roku, ale duża część dorobku nie była notowana w bazie, a jej niekompletność spowodowała podjęcie sformalizowanych działań w 2013 roku. Stąd też na podstawie odpowiedniego zarządzenia (*Zarządzenie Nr 19 Rektora Politechniki Poznańskiej z dnia 10 czerwca 2013 r. w sprawie obowiązku przekazywania do Biblioteki oraz dokumentowania i udostępniania informacji o publikacjach pracowników Politechniki Poznańskiej*) większość publikacji z roku 2013 została zaindeksowana w bazie, a kolejne będą wprowadzane na bieżąco. Poza tym również od roku 2013 na podstawie innego zarządzenia (*Zarządzenie nr 20 Rektora Politechniki Poznańskiej z dnia 10 czerwca 2013 r. (RO/VI/20/2013) w sprawie zasad gromadzenia, opracowania, udostępniania i archiwizowania utworów deponowanych w Repozytorium Naukowym Politechniki Poznańskiej*) funkcjonuje repozytorium będące źródłem publikacji pełnotekstowych (chodzi przede wszystkim o rozprawy doktorskie, habilitacyjne i podręczniki akademickie, a także artykuły i monografie Wydawnictwa Politechniki Poznańskiej). Repozytorium obejmuje obecnie 896 pełnych tekstów (dane z dnia 07.04.2014 r.).

4. Przygotowanie i import metadanych

W następnym etapie przetworzono i porównano wyeksportowane pliki w formatach BibTeX, RIS, OAI-PMH, HTML i TXT. Szczególną uwagę poświęcono formatowi BibTeX, który jest bardzo popularnym formatem do opracowywania opisów bibliograficznych i cytowania literatury i jest obsługiwany przez oprogramowanie OMEGA-PSIR. Jego zastosowanie do omawianych celów jest jednak nieco ograniczone. Po pierwsze dlatego, że standardowe typy publikacji obsługiwane przez format BibTeX nie obejmują wszystkich rozważanych typów publikacji (w tym rozpraw doktorskich i patentów). Po drugie okazało się, że relatywnie często wykrywano niewłaściwie zadeklarowane typy publikacji (szczególnie w przypadku materiałów konferencyjnych), co skutkowało nieprawidłowościami po zaimportowaniu opisów publikacji do systemu.

Ponadto analiza wykazała, że zakresy pól opisujących metadane tego samego typu publikacji, pobrane w różnych formatach z różnych źródeł informacji, różnią się między sobą i konwersja jednego formatu do drugiego jest procesem bardziej złożonym niż pierwotnie zakładano. Wymagane jest porównanie wszystkich wyeksportowanych pól i ich ponowna analiza. W takich przypadkach przygotowanie odpowiedniego skryptu do pełnego i bezbłędneho procesu konwersji jest konieczne. Najlepszym rozwiązaniem wydaje się być przygotowanie zestawu konwerterów poszczególnych formatów do natywnego formatu XML zgodnie z przyjętą definicją typów dokumentów (XSD).

Plik XML do pełnego importu danych zawiera kompleksową informację o wszystkich obiektach oraz ich hierarchii (np. artykuł, czasopismo, pracownik afiliacja) (rys. 2). Należy zwrócić szczególną uwagę na zapewnienie spójności, poprawności i kompletności danych pobranych z różnych źródeł scalonych w pliku XML, dbając o zachowanie odpowiedniej hierarchii obiektów oraz powiązań między nimi (przy czym nie chodzi o powiązania w sensie relacji, ale o pobranie danych z jednego obiektu i wstawienie ich do innego obiektu). W związku z tym bardzo ważne jest właściwe zidentyfikowanie poszczególnych obiektów nadrzędnych.

```

<?xml version='1.0' encoding='UTF-8' standalone='yes'?>
<collection>
  <article>
    <journalissue>
      <journalseries>
        <fullName>Discrete Applied Mathematics</fullName>
        <govName>DISCRETE APPLIED MATHEMATICS</govName>
        <isn>0166-218X</isn>
        <type>journalseries</type>
      </journalseries>
      <vol>10</vol>
      <no>3</no>
      <issueDate>1995</issueDate>
    </journalissue>
    <author>
      <active>true</active>
      <status>PRACOWNIK</status>
      <name>Jerzy</name>
      <surname>Namrocki</surname>
      <email>jerzy.namrocki@put.poznan.pl</email>
      <affiliation>
        <namePL>Instytut Informatyki</namePL>
        <nameEN>Institute of Computing Science</nameEN>
        <affiliation>
          <affiliationowner>PP</affiliationowner>
          <namePL>Wydział Informatyki</namePL>
          <nameEN>Faculty of Computing</nameEN>
          <affiliation>
            <affiliationowner>PP</affiliationowner>
            <namePL>Politechnika Poznańska</namePL>
            <nameEN>Poznan University of Technology</nameEN>
          </affiliation>
        </affiliation>
      </affiliation>
      <unit>iiz4</unit>
    </author>
    <title>DYNAMIC STORAGE-ALLOCATION WITH LIMITED COMPACTION - COMPLEXITY AND SOME PRACTICAL IMPLICATIONS</title>
    <doi>10.1016/0166-218X(85)90046-0</doi>
    <collation>241-253</collation>
  </article>
</collection>

```

Rys. 2. Przykładowy plik XML z danymi do importu opisu artykułu (wybrane pola)

System OMEGA-PSIR jest wyposażony w mechanizm sprawdzania duplikatów na etapie importowania. W przypadku wątpliwości wymusza na użytkowniku podjęcie określonych działań – wybór właściwego obiektu lub jego pominięcie. W wyniku przeprowadzonych testów wydaje się jednak, że lepszą metodą postępowania jest wstępne przygotowanie danych do importu w taki sposób, aby nie zawierały one duplikatów. W tym celu wykorzystano narzędzia bibliograficzne (menedżery bibliografii): EndNote, Mendeley i Zotero. Mają one bardziej rozbudowane mechanizmy usuwania duplikatów i pozwalają również na scalanie obiektów. System OMEGA-PSIR jest również wyposażony w mechanizm scalania obiektów, ale niestety nie można go użyć na etapie importu.

5. Wybrane typy publikacji

W artykule szczegółowo zaprezentowano wybrane typy publikacji, które w trakcie prowadzonych testów okazały się najbardziej problematyczne i wymagały odrębnego analizowania i podejścia.

Rozprawy doktorskie bronione na Politechnice Poznańskiej dotychczas były ewidencjonowane w katalogu bibliotecznym w wersji drukowanej od roku 1960, a od roku 2013 ich wersje elektroniczne dostępne są w Repozytorium Naukowym Politechniki Poznańskiej. Na chwilę obecną w katalogu bibliotecznym online dostępnych jest 2400 opisów bibliograficznych, a w repozytorium 45 rozpraw w formie pełnotekstowej (stan na dzień 14 czerwca 2014 r.). Rozprawy doktorskie są ważnym elementem dorobku naukowego ze względu na ich wartość naukową oraz ze względu na wymagania MNiSW dotyczące ich dokumentowania (*Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 22 września 2011 r. w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodach doktorskich, w postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora*). W związku z powyższym, wymagają zindywidualizowanego podejścia do ich ewidencjonowania,

a opracowanie szczegółowe rekordu jest bardziej złożone między innymi z tego powodu, że posiadają zdecydowanie bardziej rozbudowany opis w stosunku do pozostałych typów publikacji. Ponadto format BibTeX (o czym już wspomniano) nie spełnia oczekiwań w stosunku do rozpraw doktorskich, ponieważ nie należą one do standardowych typów publikacji obsługiwanych przez ten format.

W wyniku analizy zestawów pól metadanych opisujących publikacje w systemach: biblioteczno-informacyjnym (Horizon) i repozytorium (dLibra) oraz w przygotowywanym i wdrażanym systemie wskazano brakujące pola i odpowiednio zmodyfikowano zestaw tych pól (rys. 1, rys. 3). Ponadto zdecydowano, że optymalnym rozwiązaniem będzie zastosowanie konwersji danych do formatu XML w celu zaimportowania ich do systemu docelowego.

*Słowa kluczowe w języku polskim	przetwornik piezoelektryczny ; histereza ; kompensacja
*Słowa kluczowe w języku angielskim	piezoelectric transducer ; hysteresis ; compensation
Klasyfikacja PKT	
Klasyfikacja KBN	budowa i eksploatacja maszyn
Klasyfikacja europejska	Inżynieria mechaniczna
Pola dodatkowe	
Odzyskany tytuł: doktor nauk technicznych w dyscyplinie: budowa i eksploatacja maszyn Rozwiniędyj	
Licencja: udostępnienie na podstawie umowy z właścicielem majątkowych praw autorskich Rozwiniędyj	
Właściciel praw: Felic. Marcin Tadeusz Rozwiniędyj	
Lokalizacja oryginału: SPP Rozwiniędyj	
Prawa do dysponowania publikacją: Politechnika Poznańska Rozwiniędyj	
Prawa: wszystkie prawa zastrzeżone Rozwiniędyj	
Prawa dostępu: dla wszystkich w zakresie dozwolonego użytku Rozwiniędyj	

Rys. 3. Rozszerzanie struktury danych za pomocą pól dodatkowych – widok WWW

Patenty jako przedmioty własności intelektualnej i przemysłowej stanowią cenny dorobek naukowy uczelni technicznych, szczególnie ze względu na ich aspekt wdrożeniowy, dlatego tak ważnym elementem jest zebranie pełnej informacji (o zgłoszeniach patentowych, wynalazkach, znakach towarowych i in.) rozproszonej w bazach Urzędu Patentowego Rzeczypospolitej Polskiej (<http://www.uprp.pl/>), czy w różnych bazach międzynarodowych:

- European Patent Register - <https://register.epo.org/espacenet/regviewer>,
- Espacenet - <http://worldwide.espacenet.com>,
- PATENTSCOPE - <http://www.wipo.int/patentscope/en>
- USPTO - <http://patft.uspto.gov/>,
- Free Patents Online - www.freepatentsonline.com/,
- Google Patent Search - www.google.com/patents/,
- SciFinder <https://scifinder.cas.org/>,
- Ei Patents - <http://www.engineeringvillage.com/>.

W celu bieżącego pozyskiwania informacji o powyższym dorobku i jego wdrożeniach, a także uzupełniania go o potrzebne dane z w/w stron podjęto współpracę z Centrum Innowacji, Rozwoju i Transferu Technologii (CIRiT) Politechniki Poznańskiej. Przeanalizowano mechanizmy eksportu danych z różnych platform, gdzie możliwy jest dostęp do milionów dokumentów patentowych, ale niewiele z nich pozwala na eksport metadanych do menedżerów bibliografii. Nawet jeśli platforma wspiera takie narzędzia bibliograficzne, to wynik eksportu nie jest zadowalający. Eksport z wymienionych źródeł możliwy jest w różnych formatach: BibTeX, RIS, czy HTML, co w porównaniu do wdrażanego systemu OMEGA-PSIR, który obsługuje wyłącznie format BibTeX,

wymaga zastosowania odpowiednich narzędzi (konwerterów) do formatu BibTeX lub XML w celu pełnego importu danych i staje się dosyć skomplikowane oraz trudne do zautomatyzowania. Format BibTeX nie spełnia jednakże oczekiwań w stosunku do tego typu publikacji, o czym już wspomniano.

Niektóre menedżery bibliografii, tak jak WizFolio, wspierają bezpośredni eksport z FreePatentsOnline, Google Patents, Scirus i USPTO (do lutego 2012 r. było to również możliwe w przypadku Espacenet i SciFinder). Bardzo dobre wyniki zostały uzyskane z FreePatentsOnline i USPTO. Natomiast w przypadku zastosowania Zotero uzyskuje się słabe rezultaty jeśli chodzi o mechanizmy eksportu danych z baz patentowych.

Ponadto Mendeley i WizFolio sprawdzają się doskonale do wyciągania danych bibliograficznych zawartych w metadanych pliku PDF. Jednak ten mechanizm nie jest przydatny w przypadku patentów, ponieważ te pliki PDF nie zawierają potrzebnych metadanych.

Platformy takie jak Patent Lens (<http://www.lens.org/lens/>) i SciFinder (<https://scifinder.cas.org/>) pozwalają na eksport bibliografii patentowej w formacie RIS, który jest standardowym formatem cytatu obsługiwanym przez większość menedżerów bibliografii. Jednakże Patent Lens pozwala tylko na eksport pojedynczych rekordów. Ponadto przy eksporcie z SciFindera lepszym wyborem niż RIS jest *tagged format*, ponieważ zawiera więcej pól danych. Porównanie różnych menedżerów bibliografii, które wspierają patentowe bazy danych jest szerzej opisane w literaturze¹.

Podsumowując wyniki przeprowadzonych testów stwierdzono, że w praktyce pozostaje wykorzystanie natywnego formatu XML do importowania danych patentowych do systemu.

Materiały konferencyjne przeanalizowano na przykładach eksportu metadanych z baz: Web of Science i Scopus. Przykłady dotyczące opisów bibliograficznych wyeksportowanych w formacie BibTeX pokazują, że relatywnie często wykrywano niewłaściwie zadeklarowane typy publikacji, co wymaga przeprowadzania dodatkowych konwersji przed zaimportowaniem danych do systemu.

To pokazuje, że uporządkowanie danych przed importem według typów publikacji i pobieranie tylko określonych typów (w tym wypadku materiałów konferencyjnych), a następnie weryfikacja pobranych danych i ewentualna konwersja, pozwala na właściwe ich zaimportowanie i uniknięcie wielu błędów. Ponadto ważne jest, aby przed zaimportowaniem do docelowego systemu zawsze zweryfikować, czy źródło jest już zaindeksowane (w tym przypadku: tytuł materiałów konferencyjnych lub nazwa konferencji).

Książki, rozdziały z książek i artykuły to rodzaje publikacji, które w formacie BibTeX traktowane są jako standardowe typy. Dlatego w ich przypadku wykonywanie importu z zastosowaniem tego formatu przebiega z reguły bezproblemowo. Nie ma konieczności wykonywania konwersji do formatu XML. Niezmiennie jednak wymagana jest dbałość o zachowanie spójności i jednoznaczności danych.

Projekty jako specyficzny typ dorobku wymagają zastosowania odrębnego podejścia: opracowywania i wprowadzania na bieżąco danych o projektach. Wynika to z faktu, że informacja ta jest rozproszona na uczelni i pozyskiwanie jej będzie możliwe jedynie w kontaktach z autorami projektów lub jednostkami je koordynującymi. Stąd też na dzień dzisiejszy nie podjęto próby zautomatyzowania uzupełniania danych o projektach w systemie bazy wiedzy.

6. Podsumowanie i wnioski

Zakładano, że uda się w jednolity sposób pobrać dane z różnych źródeł, scalić je i przygotować pliki do importu w sposób kompleksowy. Rzeczywistość zweryfikowała te założenia ze względu na różnorodność platform, formatów oraz zróżnicowany stopień szczegółowości opracowania metadanych. Zautomatyzowanie całego procesu okazało się być problemem bardziej złożonym,

¹ Engineering & Science Library. Importing Patent Citations into Reference Managers [online]. [dostęp 14 czerwca 2014]. Dostępny w Internecie: <http://library.queensu.ca/research/guide/patents/download-patents/citation-managers>

wymagającym zastosowania odrębnych mechanizmów, w zależności od typów publikacji i specyfiki dorobku naukowego.

Na dalszym etapie prac przewiduje się działania dwutorowe:

1) przygotowanie zestawu narzędzi programistycznych ułatwiających osiągnięcie maksymalnej automatyzacji opisanych procesów mając na uwadze, że czynnik ludzki jest mimo wszystko niezbędny oraz

2) wypracowanie odpowiedniej organizacji pracy na poszczególnych etapach.

Sfinalizowanie projektu poprzez utworzenie centralnego repozytorium wiedzy Politechniki Poznańskiej i objęcie jednym globalnym systemem wyszukiwawczym informacji (zarówno bibliograficznej, jak i pełnotekstowej) o całym dorobku naukowym uczelni i jej pracownikach (jako ekspertach w określonych ściśle dziedzinach nauki), pozwoli na jednolite zasady jego dokumentowania, wyszukiwania i udostępniania zgodnie z prawem własności intelektualnej i prawem autorskim. Ponadto system ten pozwoli na raportowanie osiągnięć naukowych na potrzeby indywidualnej oceny pracowników, wydziałów i uczelni (mając na uwadze wytyczne MNiSW: punktacja z list A, B, C, Impact Factor) oraz współdzielenie danych z innymi systemami takimi jak PBN, czy POL-ON.

Bibliografia:

[1] Uczelniana Baza Wiedzy. Podręcznik Administratora. Wersja 1.0. Wydział Elektroniki i Technik Informacyjnych. Politechnika Warszawska. [online]. [dostęp 20 maja 2014]. Dostępny w Internecie: http://omegapsir.ii.pw.edu.pl/download/docs/dokumentacja_systemowa/Repozytorium%20PW%20-%20Podrecznik%20Administratora.pdf

[2] Uczelniana Baza Wiedzy. Instalacja i konfiguracja systemu. Wersja 1.9. Wydział Elektroniki i Technik Informacyjnych. Politechnika Warszawska. [online]. [dostęp 20 maja 2014]. Dostępny w Internecie: http://omegapsir.ii.pw.edu.pl/download/docs/dokumentacja_systemowa/Repozytorium%20PW%20-%20Instalacja%20i%20konfiguracja%20systemu%20v1_9.pdf

[3] Mau S.: Cite Patents with Bibtex. [online]. [dostęp 14 czerwca 2014]. Dostępny w Internecie: <http://www.see-out.com/sandramau/bibpat.html>

[4] Your BibTex resource. [online]. [dostęp 14 czerwca 2014]. Dostępny w Internecie: <http://www.bibtex.org/>

[5] Mendeley. Mendeley is a free reference manager. [online]. [dostęp 20 maja 2014]. Dostępny w Internecie: <http://www.mendeley.com/features/reference-manager/>

[6] Zotero. Trzy ważne informacje dla polskich użytkowników Zotero. [online]. [dostęp 24 maja 2014]. Dostępny w Internecie: https://www.zotero.org/support/pl/quick_start_guide